4E

УДК 004.89, 004.93

Т.В. Ермоленко[1], Н.С. Клименко[2]
[1]Институт проблем искусственного интеллекта МОН Украины и НАН Украины,
г. Донецк, Украина
[2]Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина
Naturewild71@gmail.com

Влияние GSM-сжатия на эффективность работы методов выделения формант

В статье описываются результаты исследований устойчивости методов выделения формант к сжатию при помощи алгоритма GSM 6.10, используемого современной сотовой связью. В работе приведен обзор ряда методов выделения формант речевого сигнала, используемых в современных системах идентификации диктора, а также численное исследование устойчивости результатов работы этих методов к сжатию с потерями.

Введение

Для получения индивидуальных параметров голоса диктора применяются различные методы, однако формантный анализ позволяет получить наиболее робастные идентификационные характеристики. Эмпирически доказано, что для характеристик звуков речи достаточно выделения четырех формант. В большинстве случаев для различения гласных звуков достаточно первых двух формант, однако практически всегда количество формант в спектре звука больше двух, что указывает на более сложные связи между артикуляцией и акустическими характеристиками звука, чем при условии рассмотрения только двух первых формант. Именно третья и четвертая форманты дают представления об индивидуальных особенностях произношения диктора, так как фиксируют побочные резонирующие частоты. Форманты в совокупности с другими характеристиками речевого сигнала представляют собой качественную динамическую оценку диктора [1], [2].

Повсеместное использование стандарта кодирования речевого сигнала GSM 6.10 обуславливает использование устойчивых методов анализа речевого сигнала в системах распознавания дикторов, анализирующих непосредственно аудиопоток мобильной связи. Подобные системы внедряются в первую очередь для применения в криминалистике при исследовании образцов речи и фонограмм разговоров в качестве вещественных доказательств.

Проблеме выделения формант в речевом сигнале посвящено много работ. Так, в [2-5] описаны результаты анализа эффективности работы используемых методов как в отдельности, так и в сравнении. Эксперименты проводились над искусственно синтезированным речевым сигналом, записями дикторов с шумовыми искажениями и выполнялись проверки устойчивости относительно типа микрофона. Тем не менее, работы, посвященные влиянию сжатия на поведение формантных траекторий, на сегодняшний день практически отсутствуют.

Данная работа посвящена актуальной в рамках фоноскопических исследований задаче – анализу влияния GSM-сжатия на эффективность ряда методов выделения формант.

Цель данной статьи – оценить влияние алгоритма GSM-сжатия на работу методов выделения формант вокализованных участков речевого сигнала. Для достижения поставленной цели были программно реализованы два метода выделения формант на основе линейного предсказания, которые отличаются концепцией выделения самих коэффициентов линейного предсказания, а также, в качестве альтернативы, рассмотрен метод нулей сигнала. Проведен сравнительный анализ устойчивости результатов работы этих методов к GSM-сжатию.

Особенности GSM-кодирования

В стандарте GSM используется метод RPE-LTP (Regular Pulse Excited Long Term Predictor – линейное предсказание с возбуждением регулярной последовательностью импульсов и долговременным предсказателем). Обработка речи осуществляется в рамках принятой системы прерывистой передачи речи (DTX), обеспечивающей включение передатчика только тогда, когда пользователь начинает разговор и отключает его в паузах и в конце разговора. DTX управляется детектором активности речи (VAD), который обеспечивает обнаружение и выделение интервалов передачи речи с шумом и шума без речи даже в тех случаях, когда уровень шума соизмерим с уровнем речи. В состав системы прерывистой передачи речи входит также устройство формирования комфортного шума, который включается и прослушивается в паузах речи, когда передатчик отключен. Экспериментально показано, что отключение фонового шума на выходе приемника в паузах при отключении передатчика раздражает абонента и снижает разборчивость речи, поэтому применение комфортного шума в паузах считается необходимым.

Кратковременное предсказание (STP – Short-Term Prediction) не обеспечивает достаточной степени устранения избыточности речи. Поэтому в дополнение к кратковременному предсказанию используется еще долговременное предсказание (LTP – Long-Term Prediction), в значительной мере устраняющее остаточную избыточность и приближающее остаток предсказания по своим статистическим характеристикам к белому шуму [6].

Формирование комфортного шума осуществляется в паузах активной речи и управляется речевым декодером. Когда VAD в передатчике обнаружит, что говорящий прекращает разговор, передатчик остается еще включенным в течение следующих пяти речевых кадров. Во время первых четырех из них характеристики фонового шума оцениваются путем усреднения коэффициента усиления и коэффициентов фильтра кодирования с линейным предсказанием (КЛП). Эти усредненные значения передаются в следующем пятом кадре, в котором содержат информацию о комфортном шуме (SID кадр). Комфортный шум генерируется на основе КЛП-анализа SID кадра. Чтобы исключить раздражающее влияние модуляции шума, комфортный шум должен соответствовать по амплитуде и спектру реальному фоновому шуму в месте передачи.

В условиях замираний сигналов в подвижной связи речевые фрагменты могут подвергаться значительным искажениям. При этом для исключения раздражающего эффекта при воспроизведении необходимо осуществлять экстраполяцию речевого кадра. Потеря одного речевого кадра может быть значительно компенсирована путем повторения предыдущего фрагмента. При значительных по продолжительности перерывах в связи предыдущий фрагмент больше не повторяется, и сигнал на выходе речевого декодера постепенно заглушается, чтобы указать пользователю на разрушение канала. То же самое происходит и с SID кадром.

Обзор методов, используемых для выделения формант

При параметризации РС с помощью КЛП используют упрощенную модель речеобразования, основанную на предположении, что РС – результат свертки сигнала возбуждения последовательностью импульсов либо случайным шумом и импульсной характеристики линейного фильтра с медленно изменяющимися параметрами, представляющего собой голосовой тракт.

Общий спектр РС, обусловленный излучением, речевым трактом и возбуждением, описывается с помощью линейной системы с переменными параметрами и моделируется как авторегрессионный процесс. Линейный предсказатель порядка p с коэффициентами a_k для сигнала $s(n)$ определяется как система, на выходе которой выполняется равенство

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) .$$

Основным подходом к получению коэффициентов предсказателя является определение параметров предсказания таким образом, чтобы минимизировать средний квадрат погрешности предсказания

$$E = \sum_{m} \left(s(m) - \sum_{k=1}^{p} a_k s(m-k) \right)^2 .$$

Это приводит к системе из p линейных уравнений с p неизвестными. Если речевой сигнал на этом интервале считается стационарным случайным процессом (автокорреляционный метод оценки коэффициентов предсказания), то система решается с помощью итерационной процедуры алгоритма Левинсона-Дарбина [1]. Если речевой сигнал считается нестационарным процессом (ковариационный метод оценки коэффициентов предсказания), система решается с использованием разложения Холецкого [5].

После получения моментального спектра РС, вычисленного по КЛП, находятся его пики либо, в их отсутствие, центры плотности. Частоты, соответствующие этим пикам, и будут представлять собой формантные частоты.

Метод нулей сигнала для оценки формантных частот [3], [4] основан на анализе распределения длительностей интервалов между нулями сигнала. Анализ нулей сигнала предполагает, что в данной частотной полосе присутствуют колебания только одной форманты. Это связано с известным свойством, согласно которому при наличии нескольких частот средняя частота переходов определяется как средневзвешенная по амплитудам каждой частоты. Именно поэтому в методе нулей сигнала особенно важен выбор полос частот для анализа.

Исследование эффективности методов

Для проверки устойчивости вышеописанных методов вычисления формант к GSM-сжатию они были реализованы в едином программном комплексе, после чего проводилось численное исследование эффективности их работы.

Тестирование методов проводилось на звуковых сигналах одного диктора, записанных в формате WAV PCM с частотой дискретизации 44.1 кГц и глубиной квантования 16 бит. Диктор произносил слова «один», «четыре», «труд». Кроме того, для каждого из слов была сделана запись в формате GSM 6.10 WAV с частотой дискретизации 8 кГц. Все записи были созданы с помощью программы Audacity 1.3.12-beta и осуществлялись в монорежиме.

Для оценки брались первые четыре форманты F1 – F4, полученные тремя различными методами по реализациям звуков [а], [и], [у] в вышеперечисленном речевом материале, и сравнивались со значением, определенным экспертом по спектрограммам звуков (табл. 1).

Таблица 1 – Значения формантных частот, определенных экспертом

Звук	F1, Гц	F2, Гц	F3, Гц	F4, Гц
А	688	1505	2236	3483
И	301	2107	3010	3913
У	344	860	2323	3354

В табл. 2 – 4 приведены результаты численного исследования эффективности методов на записанном речевом материале. Математическое ожидание и среднее квадратичное отклонение (СКО) каждой из формант считалось по временному ряду, построенному следующим образом: речевой сигнал разбивался фреймами длиной 1024 отсчета с половинным перекрытием, на каждом из фреймов, входящих в соответствующий звук, вышеописанными методами вычислялись значения F1 – F4, из которых и формировались 4 выборки, соответствующие формантным частотам. Погрешность вычислялась как модуль разности между математическим ожиданием и истинным значением форманты, определенным экспертом, значение ошибки представляет собой процентное отношение погрешности к истинному значению форманты.

В силу специфики метода нулей необходимо указание особых частотных полос для анализа каждого звука. Однако таким образом метод становится сильно зависим от работы эксперта, что лишает его автономности в принятии решений. Для данного исследования были взяты статистически универсальные значения полос: 300 – 1000 Гц для F1, 1000 – 1800 Гц для F2,1800 – 2500Гц для F3 и 2500 – 4000 Гц для F4.

Результаты исследования устойчивости метода автокорреляционных КЛП для звуков [а], [и], [у] приведены в табл. 2.

Результаты исследования устойчивости метода ковариационных КЛП для звуков [а], [и], [у] приведены в табл. 3.

Результаты исследования устойчивости метода нулей сигнала для звуков [а], [и], [у] приведены в табл. 4.

По полученным результатам численных исследований можно судить о качестве работы методов и их устойчивости к GSM-сжатию. Как и следовало ожидать, погрешность оценки формант всеми методами увеличивается при работе с GSM-кодированным сигналом, что обусловлено частичной потерей и искажением речевых данных.

Наименьший разброс оценок формант показал метод нулей сигнала. Значения формант, полученные с помощью этого метода, довольно четко сконцентрированы в диапазоне предполагаемой формантной частоты. Сжатие речевого сигнала привело к увеличению погрешности в среднем не более, чем на 2%, что позволяет считать данный метод устойчивым к GSM-сжатию. Тем не менее среди всех рассмотренных методов для сигнала формата WAV метод нулей дает самую большую погрешность. Это обусловлено зависимостью этого метода от выбора статистических полос оценок формант. Практически невозможно разделить пространство частот таким образом, чтобы в каждой из них находилось не более одной форманты при рассмотрении любой фонемы.

Таблица 2 – Оценка формант гласных звуков методом автокорреляционных КЛП

Форманта	Частота, определенная экспертом, Гц	Мат. ожидание, Гц	СКО, Гц	Погрешность, Гц	Ошибка, %
Оценка формант звука «А» в сигнале формата WAV					
F1	688	705,8	57,46	17,8	2,59
F2	1505	1523,46	180,02	18,46	1,23
F3	2236	2275,69	145,57	39,69	1,78
F4	3483	3180,13	449,11	302,47	8,69
Оценка формант звука «А» в GSM-кодированном сигнале					
F1	688	715,02	68,99	27,02	3,93
F2	1505	1512,49	108,15	7,49	0,5
F3	2236	2250,4	185,55	14,4	0,64
F4	3483	3192,07	464,23	290,53	8,34
Оценка формант звука «И» в сигнале формата WAV					
F1	301	280,97	38,05	19,76	6,57
F2	2107	2196,91	94,18	89,91	4,27
F3	3010	2967,21	152,58	42,79	1,42
F4	3913	3129,78	307,44	783,22	20,02
Оценка формант звука «И» в GSM-кодированном сигнале					
F1	301	285,24	40,01	15,49	5,15
F2	2107	2186,96	113,37	79,96	3,8
F3	3010	2978,22	197,87	31,78	1,06
F4	3913	3086,4	307,88	826,6	21,12
Оценка формант звука «У» в сигнале формата WAV					
F1	344	276,81	35,55	67,19	19,53
F2	860	840,69	76,06	19,12	2,22
F3	2323	2380,32	163,79	57,77	2,49
F4	3354	3123,29	419,03	230,71	6,88
Оценка формант звука «У» в GSM-кодированном сигнале					
F1	344	282,19	45,95	61,81	17,97
F2	860	836,99	98,15	22,83	2,66
F3	2323	2387,06	235,61	64,47	2,78
F4	3354	3184,56	397,85	169,44	5,05

Таблица 3 – Оценка формант гласных звуков методом ковариационных КЛП

Форманта	Частота, определенная экспертом, Гц	Мат. ожидание, Гц	СКО, Гц	Погрешность, Гц	Ошибка, %
Оценка формант звука «А» в сигнале формата WAV					
F1	688	831,34	243,85	143,34	20,83
F2	1505	1691,1	198,46	186,1	12,37
F3	2236	1993,09	280,81	242,91	10,86
F4	3483	3273,4	223,88	209,2	6,01
Оценка формант звука «А» в GSM-кодированном сигнале					
F1	688	765,41	165,86	77,41	11,25
F2	1505	1229,51	297,76	275,49	18,3
F3	2236	2000,79	232,81	235,21	10,52
F4	3483	3460,03	277,08	22,57	0,65
Оценка формант звука «И» в сигнале формата WAV					
F1	301	425,73	18,69	125	41,56
F2	2107	2136,71	48,8	29,71	1,41
F3	3010	3209,09	71,02	199,09	6,61
F4	3913	3252,25	48,07	660,75	16,89
Оценка формант звука «И» в GSM-кодированном сигнале					
F1	301	472,31	369,2	171,58	57,05
F2	2107	2131,73	97,73	24,73	1,17
F3	3010	3055,9	212,22	45,9	1,53
F4	3913	3070,48	215,1	842	21,53
Оценка формант звука «У» в сигнале формата WAV					
F1	344	430	0	86	25
F2	860	467,28	42,62	610,54	29,65
F3	2323	1994,49	653,01	328,06	14,13
F4	3354	2411,2	654,16	942,8	28,11
Оценка формант звука «У» в GSM-кодированном сигнале					
F1	344	431,79	12,28	87,79	25,52
F2	860	481,32	60,78	378,5	44,02
F3	2323	1930,23	656,81	392,33	16,89
F4	3354	2461,97	567,22	892,03	26,6

Таблица 4 – Оценка формант гласных звуков методом нулей

Форманта	Частота, определенная экспертом, Гц	Мат. ожидание, Гц	СКО, Гц	Погрешность, Гц	Ошибка, %
Оценка формант звука «А» в сигнале формата WAV					
F1	688	652,13	37,58	35,88	5,21
F2	1505	1481,21	218,36	23,79	1,58
F3	2236	1874,81	77,13	361,19	16,15
F4	3483	2745,97	163,36	736,63	21,15
Оценка формант звука «А» в GSM-кодированном сигнале					
F1	688	680,63	66,78	7,38	1,07
F2	1505	1266,03	72,65	238,97	15,88
F3	2236	2055,22	282,12	180,78	8,08
F4	3483	2729,16	159,23	753,44	21,63
Оценка формант звука «И» в сигнале формата WAV					
F1	301	327,96	23,28	27,22	9,05
F2	2107	2010,63	81,75	96,37	4,57
F3	3010	2567,5	110,59	442,5	14,7
F4	3913	2657,62	226,46	1255,38	32,08
Оценка формант звука «И» в GSM-кодированном сигнале					
F1	301	327,19	17,27	26,46	8,8
F2	2107	2034,82	85,22	72,18	3,43
F3	3010	2606,24	256,81	403,76	13,41
F4	3913	2614,97	250,41	1298,03	33,17
Оценка формант звука «У» в сигнале формата WAV					
F1	344	534,11	261,53	281,11	18,27
F2	860	936,88	145,16	77,07	8,96
F3	2323	2193,55	137,87	129,01	5,55
F4	3354	2646,38	115,83	707,62	21,1
Оценка формант звука «У» в GSM-кодированном сигнале					
F1	344	487,46	247,75	268,46	22,7
F2	860	979,58	118,48	119,77	13,93
F3	2323	2186,02	124,96	136,54	5,88
F4	3354	2644,91	115,2	709,09	21,14

Индивидуальность голоса дикторов еще больше усложняет данную задачу. Поэтому при численном исследовании были нередки случаи попадания нескольких формант в одну полосу анализа. Это обстоятельство значительно повлияло на эффективность работы, результат которой зачастую сводится к выбору более низкой частоты или к девиантным колебаниям между формантами. Таким образом, метод требует тонкого подбора параметров, что является нежелательным при автоматическом распознавании. Следовательно, применение данного метода в формантном анализе возможно лишь в узкой области задач, требующих тесного взаимодействия с экспертом.

Оценки формант по методу ковариационных КЛП отличаются незначительным увеличением СКО по сравнению с методом нулей, форманты оцениваются довольно точно (погрешность 1 – 10%). Однако СКО увеличивается в несколько раз при обработке GSM-кодированного РС. Искажения, вносимые сжатием, в значительной мере «размывают» границы частотных пиков. Этим обусловлена нестабильность оценок формантных частот, что делает данный метод неустойчивым к GSM-сжатию.

Погрешность оценок формант по методу автокорреляционных КЛП для сигналов формата WAV является наиболее низкой из всех полученных. Особенно точно определяются частоты формант F2 и F3 (погрешность 0,5 – 4%). Разброс оценок формант GSM-кодированного РС остается практически таким же, как и для несжатого РС, что выгодно отличает данный метод от других, СКО увеличивается в значительной мере.

Исследования показали, что GSM-сжатие сильно влияет на эффективность работы метода ковариационных КЛП, в то время как методы нулей сигнала и автокорреляционных КЛП показали высокую устойчивость к данному типу кодирования РС и могут применяться для эффективного формантного анализа. Что касается точности оценок формантных частот, то эти методы показывают хорошие результаты только при выполнении своих специфических условий: метод нулей требует выделения точных полос анализа, а КЛП-методы – низкого уровня шума и искажений.

Выводы

Таким образом, было установлено, что рассмотренные методы значительно различаются между собой в показателях погрешности оценок формант, их СКО и устойчивости.

Наименьшие погрешности для несжатого сигнала достигаются при вычислении формант методом автокорреляционных КЛП, наибольшие – методом нулей, что обусловлено спецификой работы алгоритма, и могут быть снижены использованием эвристик или вмешательством эксперта. Оба метода характеризуются относительно небольшим СКО оценок формант как для сжатого, так для несжатого сигнала.

Наибольшую величину СКО оценок формант для GSM-кодированного РС дает метод ковариационных КЛП. GSM-сжатие сильно влияет на эффективность его работы, в то время как методы нулей сигнала и автокорреляционных КЛП показали высокую устойчивость к данному типу сжатия РС и могут применяться для эффективного формантного анализа.

Полученные в работе результаты могут быть использованы при разработке систем идентификации диктора в задачах фоноскопической экспертизы, в системах контроля доступа и биометрической идентификации.

Литература

1. Рабинер Л.Р. Цифровая обработка речевых сигналов / Рабинер Л.Р., Шафер Р.В. ; [пер. с англ.]. – М. : Радио и связь, 1981. – 496 с.

2. Reynolds D. Experimental evaluation of features for robust speaker identification / D. Reynolds // IEEE Trans. On Speech and Audio Processing. – 1994. – № 4, vol. 2. – 870 p.
3. Сорокин В.Н. Устойчивость оценок формантных частот / Сорокин В.Н., Леонов А.С., Макаров И.С. // Речевые технологии. – 2009. – № 1. – С. 3-21.
4. Леонов А.С. К анализу резонансных частот речевого тракта / А.С. Леонов, В.Н. Сорокин // Информационные процессы. – 2007. – № 4, т. 7. – С. 386-400.
5. Сорокин В.Н. Об автокорреляционном анализе речевых сигналов. / В.Н. Сорокин, И.П. Трифоненков // Акустический журнал. – 1996. – № 3, т. 42. – С. 368–374.
6. Иванов И.Л. Экспертное исследование формата GSM [Электронный ресурс]. – Режим доступа : http://www.illidiy.orel.ru/Pub/publ6.htm

Literatura

1. Rabiner L.R. Digital speech signalsprocessing. M.: Radio i Svyaz. 1981. 496 s.
2. Reynolds D. Experimental evaluation of features for robust speaker identification. №4. Vol 2. 1994. 870 s.
3. Sorokin V.N. Speech technologies. № 1. 2009.S 3-21
4. Leonov A.S. Kanalizuresonansnihchastotrechevogotrakta. № 4.Vol 7. 2007. S 386-400
5. Sorokin V.N. Acoustical Physics. №3, Vol 42. 1996. S 368-374
6. IvanovI.L.Ekspertnoeissledovanieformata GSM. http://www.illidiy.orel.ru/Pub/publ6.htm

Т.В. Єрмоленко, М.С. Клименко

Вплив GSM-стиснення на ефективність роботи методів виділення формант

У статті описано результати досліджень стійкості методів виділення формант до стиснення за допомогою алгоритму GSM 6.10, що використовується у сучасному стільниковому зв’язку. У статті наведено огляд ряду методів виділення формант мовленнєвого сигналу, що використовуються у сучасних системах ідентифікації диктора, а також чисельне дослідження стійкості результатів роботи цих методів до стиснення із втратами.

T.V. Yermolenko, M.S. Klymenko

Influenceof GSM-Compression on the Features of Formant Tracking Methods

The paper is devoted to the problem of formant tracking methods robustness to GSM 6.10 compression algorithm, which is employed within modern cellular networks. This article describes methods of formant tracking used in modern speaker identification systems. Computational investigation results of these methods robustness to lossy compression are also shown.

Статья поступила в редакцию 01.07.2011.